

The Challenge of Selection Bias and Confounding in Palliative Care Research

Helene Starks, Ph.D., M.P.H.,¹ Paula Diehr, Ph.D.,² and J. Randall Curtis, M.D., M.P.H.^{1,3}

Introduction

THE EARLY SUCCESSES of hospital- and community-based hospice and palliative care programs have led to rapid growth of these services over the past several decades.^{1,2} This widespread adoption is generally seen as a positive addition to the care of patients with complex illness. However, the growth in palliative care and hospice programs creates challenges for research, especially in the use of randomized controlled trial (RCT) designs, which are deemed the gold standard of research methods because of their ability to control for biases in the estimates of treatment effects. As Carlson and Morrison³ noted in their introduction to this research methods series, sometimes RCTs can be impossible to conduct due to inherent challenges in recruiting seriously ill participants, or may not be ethically appropriate in the absence of clinical equipoise. To date, observational studies have provided much of the evidence demonstrating the effectiveness of palliative care interventions on outcomes including pain and symptom management,⁴⁻⁹ communication between clinicians, patients, and families,¹⁰⁻¹³ patient and family satisfaction with care,^{8,14} and on reducing the costs of care while maintaining or improving quality.^{8,15-20} Observational studies have been criticized for their methodological weaknesses, yet it would be a mistake to dismiss their use as they have and will continue to offer important insights into the real world of hospice and palliative care.

Selection Bias and Confounding

Two of the major methodological challenges in observational research are selection bias and confounding which can contribute to underestimates or overestimates of the actual effect of an intervention (or treatment or exposure). Selection bias occurs when nonrandom factors influence enrollment into an intervention, such as referral to hospice or a palliative care consult service. Selection bias may be particularly problematic in observational studies of interventions or treatments when eligibility criteria limit entry into the intervention or because of characteristics of patients, clinicians, systems, or environments that influence the choice of who will receive the intervention. Another type of selection bias is "healthy volunteer bias" that occurs when those who participate in research or who remain in longitudinal studies are generally healthier than those who do not.

Selection bias may lead to confounding, which occurs when the set of variables that predispose selection into the intervention are also related to the outcome (Fig. 1). The association of these variables with both the intervention and the outcome can result in type I errors in which the outcomes of the intervention (known as treatment effects) are falsely attributed to the intervention rather than to the confounding variables. For example, a critique of studies demonstrating cost savings of hospice programs is that the people who choose (or are referred to) hospice are systematically different from those who do not choose hospice: the cost savings may be due to patients' personal preferences to avoid aggressive interventions and high cost environments such as the intensive care unit, rather than the hospice program. Thus, their preferences for less aggressive care may be both a reason for referral and an explanation of their lower use of high cost services. Alternatively, confounding can result in a type II error, in which the study incorrectly concludes there are no treatment effects. For example, patients referred to a palliative care consult service (PCS) may have higher symptom burdens or more comorbid conditions and organ systems requiring treatment. The PCS group may have the same or higher costs for these reasons alone, unless they are compared to similarly complicated patients who did not receive a consult. Without adequately accounting for severity of illness and baseline symptoms, palliative care could appear to be the source of the higher costs.

Figure 1 depicts the relationship between confounding variables and intervention and outcome variables. We include multiple terms for these different variables in recognition of the need to include language that is used by the many disciplines that comprise palliative care.²¹ We refer to the intervention as the independent variable; this is also known as the treatment, intervention, or exposure variable (in design terms) and the predictor, explanatory, or input variable (in analytic terms). In the figure, in-patient palliative care consults are the independent variable. In this example, the outcome variables (also known as the dependent, response, explained, or output variables) are the costs of hospital care. The confounding variables include both measured and unmeasured (or difficult to accurately measure) factors that are related to both the intervention and the outcome. Important domains that should be considered when evaluating studies include those that might lead to differential se-

¹Department of Bioethics and Humanities, School of Medicine, ²Department of Biostatistics, School of Public Health, ³Division of Pulmonary and Critical Care Medicine, Department of Medicine, School of Medicine, University of Washington, Seattle, Washington.

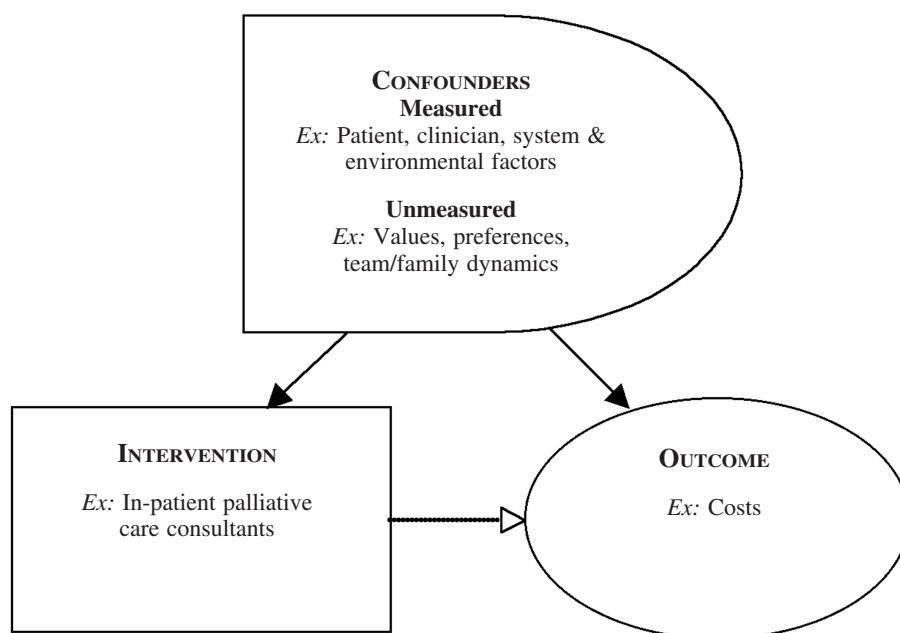


FIG. 1. Relationship between confounder, intervention, and outcome variables.

lection into hospice or palliative care programs such as severity of illness, symptom burden, functional status, prognosis, values and preferences for quality of life and life-sustaining treatments, social support, and financial resources. These variables are not always easy to capture. In the figure, examples of measured factors could include patient demographics and health status, clinician practice characteristics, type of institution and urban/rural location. Influential unmeasured factors in palliative care often include patient, family and/or clinician preferences for types of care, or differences of opinion about the goals of care or treatment options.

In observational studies, the main concerns with confounding are that (1) the potentially confounding variables may have very different distributions in the intervention and comparison groups due to selection bias and (2) estimates of treatment effects may be affected by residual confounding because of unmeasured or poorly measured variables.

Methods to Address Confounding

There are three general approaches that are used to control for confounding in observational studies: multivariable regression modeling, propensity scores, and instrumental variables. This paper addresses the first two methods; the third is the subject of the next paper in this series.²² Multivariable models using linear regression (for continuous outcomes, such as costs), logistic regression (for binary outcomes, such as mortality), or Cox proportional hazards models (for temporal outcomes, such as survival time) are the most common methods used to control for confounding. Regression models control for confounding by estimating (and specifying) the contribution of each variable to the outcome, while holding all the other variables constant in the model. The choice of variables to include in the model will depend on the question under study, sample size, and the availability of relevant variables. The objective is to include

a set of variables that are theoretically or actually correlated with both the intervention and the outcome to reduce the bias of the estimate of the treatment effect.^{23,24} Bias refers to the difference between the estimated mean value and the “true” value (which can never actually be known). Including more potential confounders in the regression may decrease the bias of the treatment effect; however, adding more variables can decrease statistical power in small samples because it increases the variance (spread) around the regression estimate by decreasing the number of degrees of freedom. Thus, the goal of model building is to carefully select the best set of confounding variables that includes the most important factors likely to account for differences between intervention and comparison groups and achieves a balance in the trade-off between bias and variance to obtain more precise estimates of the treatment effects. Readers should consider how well the included variables control for confounding and whether important variables are missing from the model.

Propensity scores also control for confounding and are similar to multivariable modeling except that an initial model is generated using variables that predict assignment to the intervention group rather than predict the outcome. This method allows investigators to examine and control for the distribution of confounders in both intervention and comparison groups by providing a summary measure of the conditional probability of being assigned to the intervention group (regardless of actual group assignment) based on a set of confounders.^{23–29} The scores range from 0 to 1; the score for a particular participant represents the estimated probability of being assigned to the intervention group, given that person’s particular combination of covariates. Participants with the same set of covariates will have the same score. Propensity scores are useful descriptively and are often stratified so that the treatment effect can be described in persons who are either least likely (scores in the lowest stratum) or

TABLE 1. CONFOUNDER VARIABLES USED IN ANALYSES EXAMINING THE EFFECTS OF HOSPICE AND PALLIATIVE CARE INTERVENTIONS ON COST

<i>Authors</i>	<i>Intervention</i>	<i>Patient variables</i>	<i>Contextual variables</i>
Campbell et al. ¹⁵	Hospice use among Medicare beneficiaries	Gender Race (white/non-white) Enrollment in Medicaid Primary condition Comorbid conditions	Urban setting (metropolitan/urbanized/rural) Nursing home residence in states with low (20–29%) vs. high (> 30%) hospice use
Gozalo et al. ¹⁶	Hospice use among nursing home residents	Age (in 3 categories > 65) Gender Race (white/black/other = mostly Hispanic) Health conditions (cancer, dementia, CHF, peripheral vascular disease, septicemia, emphysema/COPD, pneumonia, hip fracture, hypertension, pressure ulcers) Health status (change in self-sufficiency, bowel incontinence) Do-not-resuscitate order Do-not-hospitalize order High use of health care in months 7–12 months before death	Nursing home characteristics (hospital- vs. community-based, for-profit status, occupancy rate, size (≥ 120 beds)) Distance to nearest hospice Geographic region in Florida (N, central, S)
Taylor et al. ¹⁷	Hospice use among Medicare beneficiaries who died	Age at death Gender Ethnicity (Caucasian vs. non-Caucasian) Medicare expenditures in the year prior to year of death Primary health condition (cancer [lung/other], heart disease, cerebrovascular disease, COPD, nephrotic conditions, Alzheimer's, diabetes, other [infections, accidents]) Duration of primary health condition	Year of death Census region (one of 9)
Morrison et al. ¹⁸	Palliative care consultation services in 8 community & academic hospitals	Age Gender Marital status Insurance Primary diagnosis Comorbidity score	Attending physician specialty
Penrod et al. ¹⁹	Palliative care consultation services in 12 VA hospitals	Age Gender Principal diagnosis (cancer, infectious disease, cardiovascular, pulmonary, GI, genitourinary, other) Comorbidities	Hospital site Service type (medicine vs. surgery)

CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; GI, gastrointestinal.

most likely (scores in the highest stratum) to receive the intervention.

Propensity scores can be used to improve estimates of the treatment effect in three ways. First, they can be used in regression models to summarize and control for the set of confounders used to compute the score. This application produces similar results to multivariable regression and can be useful in situations with small sample sizes and limited power to detect differences between groups because it collapses the set of variables to a single score and requires only one degree of freedom for all the covariates in the model. Second, they can be used to find stratum- or individual-level matches from the comparison group whose propensity scores are similar to those in the intervention group. Third,

they can be used as a population weight to account for the distribution of included variables. In the context of palliative care, propensity scores have been used to examine the effects of a variety of interventions including hospice, hospital-based consult services, and home-based care management programs. Possible outcomes include costs and resource utilization, symptom management, quality of care, and patient and family satisfaction with care.^{15–20,30–36}

Variable Selection in Multivariable Modeling and Propensity Score Analyses

To illustrate how these methods are used, we reviewed five studies that used multivariable modeling and propen-

TABLE 2. PARTICIPANT CHARACTERISTICS BEFORE AND AFTER MATCHING ON PROPENSITY SCORE

	<i>All controls n = 14,249</i>	<i>PCS n = 319</i>	<i>p value</i>	<i>PCS n = 319</i>	<i>Matched controls n = 635</i>	<i>p value</i>
Age, mean (SD)	45.2 (19.3)	63.5 (17.9)	<0.001	63.5 (17.9)	64.5 (17.6)	0.39
Comorbidities, n (SD)	0.6 (1.2)	2.6 (2.4)	<0.001	2.6 (2.4)	2.5 (2.4)	0.32
Male	64.4%	61.1%	0.23	61.1%	64.1%	0.37
Cancer	2.2%	14.1%	<0.001	14.1%	18.0%	0.13
Metastatic cancer	0.9%	20.4%	<0.001	20.4%	15.6%	0.07
COPD	11.1%	22.3%	<0.001	22.3%	21.6%	0.81
Cerebrovascular	8.7%	21.9%	<0.001	21.9%	21.7%	0.94
CHF	4.8%	19.1%	<0.001	19.1%	21.4%	0.41
Renal	3.0%	13.2%	<0.001	13.2%	10.7%	0.26
AMI	4.4%	10.3%	<0.001	10.3%	13.1%	0.22
Hemi/paraplegia	1.4%	4.4%	<0.001	4.4%	3.9%	0.74

PCS, Palliative Care Service patients who received at least one consult; COPD, chronic obstructive pulmonary disease; CHF, congestive heart failure; AMI, acute myocardial infarction; SD, standard deviation.

sity scores to examine the effect of hospice services or palliative care consultations on costs.^{15–19} These studies used administrative data from either Medicare, Medicaid, VA, or hospital cost accounting sources. Administrative data are frequently used in observational studies because they are readily available, cover large populations (which increases sample size), and allow for comparisons across multiple institutions and geographic regions. However, they often lack specific patient, clinician, system, or environmental variables that might represent important sources of residual confounding (the amount of bias in the estimated treatment effects resulting from unmeasured or poorly measured confounders). Table 1 summarizes the patient and contextual variables used in these five studies. These variables meet the criteria of being possible confounders because they are theoretically associated with both the intervention (hospice or palliative care consultation) and the outcome (costs). We list them as a guide to variable selection for future studies and to help evaluate the variables used in analyses of administrative data. The patient variables included demographics (age, gender, race, marital status), primary and comorbid health conditions, and measures of functional status. The contextual variables included clinician factors (attending physician specialty), and institutional and environmental factors including where the intervention was delivered (hospital, service [medicine versus surgery], nursing home), geographic location, proximity to hospices, and for-profit status. A shortcoming of using administrative data for palliative care is the absence of key confounding variables such as pa-

tient/clinician treatment preferences or goals of care. Thus while there are many good reasons to use these data, readers should evaluate results carefully, especially when treatment effects are small.

Choice of Method to Control for Confounding

In three systematic reviews of the literature comparing multivariable regression to propensity score analyses, the authors all conclude that propensity scores add little to improving the estimates of treatment effects when compared with multivariable regression models.^{37–39} For example, Stürmer and colleagues³⁸ found only 9 (of 69) studies in which propensity scores contributed a greater than 20% difference in estimates of treatment effects, compared to regression. The study by Penrod and colleagues provides an illustration of this finding.¹⁹ They estimated cost differences between patients who did and did not receive a palliative care consult using an unadjusted model, a multivariable model controlling for confounders, and a model using propensity scores. They demonstrated significant savings for the group that received the palliative care consult in terms of average daily total direct costs and daily ancillary costs. Differences in the estimates from the three models were \$397, \$239, and \$234 for daily total direct costs, and \$102, \$98, and \$85 for daily ancillary costs, respectively.

The modest differences in estimating treatment effects between the multivariable regression and propensity score models invites the question about why one should bother to

TABLE 3. STRATIFICATION BY PROPENSITY SCORE

	<i>Quintiles</i>					<i>Total</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	
PCS cases, n (%)	3 (0.9)	11 (3.4)	29 (9.1)	43 (13.5)	233 (73.0)	319 (100)
Non-PCS cases, n (%)	2,926 (20.5)	2,927 (20.5)	2,848 (20.0)	2,870 (20.1)	2,678 (18.8)	14,249 (100)
Total hospital days PCS, mean	10.7	24.1	19.2	34.2	17.7	20.2
Total hospital days, non-PCS, mean	5.3	8.2	9.0	10.5	10.6	8.7
Difference between groups	5.4	15.9	10.2	23.6	7.1	11.5

PCS, Palliative Care Service.

TABLE 4. STRENGTHS AND LIMITATIONS OF METHODS TO ADJUST FOR CONFOUNDING IN OBSERVATIONAL STUDIES

	<i>Strengths</i>	<i>Limitations</i>
Multivariable modeling	Most commonly used method, many people familiar with techniques and assumptions Allows examination of the effect of individual confounders on outcome	Cannot detect group differences in distributions of confounders Model assumptions may not fit the data Only adjusts for measured confounders
Propensity score matching	Useful with small datasets as it summarizes a set of confounders into a single variable Increases internal validity—shows differences for patient groups with similar characteristics Allows for closer examination of distributions of confounders and outcomes across strata	Balances confounders at the group level; similar scores may not represent the same mix of confounders at the individual level Only adjusts for measured confounders

use propensity scores, especially when most readers are already familiar with regression analyses. The answer depends on the purpose of the analysis. A limitation of multivariable regression is that it does not provide any specific information about the degree of imbalance in the distribution of measured confounders between the intervention and comparison groups. Propensity scores are primarily used to address this limitation. Most studies use propensity scores for either matching or stratification. Matching is useful when the goal is to demonstrate efficacy, as it increases the internal validity of the estimates by making comparisons between similar groups. Matching on propensity scores has the advantage of evenly distributing the effects of a large number of confounders across the intervention and control groups. However, unlike a RCT, propensity scores cannot match on or control for unobserved confounders.

It is nearly impossible to find matches on a large number of individual variables, but it is possible to find a good match on the propensity score, which is a summary of those variables. Matching is typically done either at the individual level or by stratum, depending on the sample size, availability of potential matches and what comparisons make sense for the outcomes of interest. Frequently, observational studies using administrative data have large, heterogeneous samples. If the goal is to conduct a case-control study, then matching at the individual level is appropriate and can be done using a variety of methods, each of which estimates the distance between the scores of intervention patients and comparison patients. Specific methods include different distance measures (nearest neighbor, radius, kernel or Mahalanobis) as well as decisions about the number of matches (pairwise, 2-to-1, or multiple matching with and without replacement).^{40,41} Alternatively, a subset of the dataset can be matched on strata that represent the cases and controls that have the highest probability of meeting the entry criteria for being a case. However, if the goal is to generalize results to the entire population, matching by stratum may permit the use of all the data and provides additional information about the distribution of confounders across the entire sample.

After computing propensity scores, investigators examine how well the matching actually distributed the measured confounders in both groups, much as one would do in an RCT to check that the randomization process worked. Authors do not always have room to report the differences in

distributions of confounders before and after matching on propensity scores. We include an example in which we computed propensity scores to select matches for patients who received at least one consultation from a hospital-based palliative care service (PCS). Table 2 reports the comparisons of patient characteristics between PCS patients and the total population of hospitalized patients (on the left) and the subsample of patients who were matched 2-to-1 using the propensity scores (on the right). Nearly all the confounders had significant differences when comparing the total sample to the PCS patients, yet those differences disappeared for all the variables in the matched sample. The absence of significant differences between the matched groups indicates that the measured confounders are evenly balanced; however, one must remember that this provides no assurance that the distributions of unmeasured confounders are balanced. These results are typical and should be expected in studies using propensity scores for matching.

When propensity scores are used for stratification, they are computed for each individual and then the entire sample is evenly divided into strata; typically five strata (quintiles) are used as this number has been shown to account for the majority of bias due to imbalance on the covariates.^{26,27,42} Comparisons can then be made between the intervention and comparison groups across strata. Examining the distribution of the covariates within the quintiles can be useful to understand where the covariates are balanced across the sample. Failure to achieve balance suggests it may not be possible to make valid comparisons between treatment and control in some of the strata. This information may explain choices about which participants are included in the analysis; restricting the sample to those most likely to receive the intervention increases the validity and generalizability of the findings. Table 3 reports another example from our data to illustrate how propensity scores can help examine differential effects across the strata. We divided the sample into quintiles and report the distribution of patients in the two groups across the five strata. We also computed the total number of hospital days for each patient, summed across all hospitalizations, then estimated the mean for the PCS and comparison patients.

There are a few key points to notice about Table 3 that could guide interpretation of these data. First is the sample size within quintiles for the intervention and comparison

groups. By definition, the propensity score computes the probability of being in the intervention group, thus it follows that patients in the PCS group are extremely unlikely to be in quintile 1 (<1%, $n = 3$) and highly likely to be in quintile 5 (73%, $n = 233$). Examining the distributions, one may decide to restrict analysis to only those persons in quintile 5 to generalize results from the study population to "typical" patients receiving palliative care interventions. The differences in total hospital days illustrate this point: the greatest difference is for PCS patients who are "atypical" yet receive a consult (quintiles 1–4). For "typical" PCS patients (quintile 5), the difference is relatively small. In this example, the difference of 7.1 days more hospital days for patients receiving a consult would be generalizable to all persons who meet the criteria for quintile 5. As mentioned above, one-to-one matching is another choice, with results generalizable to yet another population.

Conclusion

The use of either multivariable regression or propensity scores in observational studies can increase the validity of these studies and the precision of estimates of the treatment effects. As the field of palliative care grows, there is much to learn about variations in practice and how interventions work in different settings and with different patient populations. These methods increase the utility of existing data, which facilitates opportunities for regional, national, and international collaborations to examine hospice and palliative care outcomes.

Table 4 summarizes the strengths and limitations of multivariable modeling and propensity scores in controlling for confounding. The major limitation of both these methods is that they can only account for what can be accurately measured; they have no direct influence on controlling the bias due to unmeasured confounders, except to the extent that those factors are highly correlated with measured characteristics. However, these methods can reduce the error associated with measured characteristics and increase the methodological rigor of observational studies. These statistical methods are important because observational data sources and study designs will continue to be strong contributors to building the evidence base for palliative care research and practice.

Acknowledgments

This work was supported by a career development award from the National Palliative Care Research Center for Dr. Starks, and a midcareer investigator award from the National Heart Lung and Blood Institute (K24-HL-068593) for Dr. Curtis.

References

- Morrison RS, Maroney-Galin C, Kralovec PD, Meier DE: The growth of palliative care programs in United States hospitals. *J Palliat Med* 2005;8:1127–1134.
- National Hospice and Palliative Care Organization. NHPCO facts and figures: Hospice care in America. www.nhpc.org/files/public/Statistics_Research/NHPCO_facts-and-figures_Nov2007.pdf (Last accessed September 19, 2008).
- Carlson MDA, Morrison RS: User's guide to research in palliative care: Why is a new series needed? *J Palliat Med* 2008;11:1258–1261.
- Finlay IG, Higginson IJ, Goodwin DM, Cook AM, Edwards AG, Hood K, Douglas HR, Normand CE: Palliative care in hospital, hospice, at home: Results from a systematic review. *Ann Oncol* 2002;13(Suppl 4):257–264.
- Grande GE, Todd CJ, Barclay SI, Farquhar MC: A randomized controlled trial of a hospital at home service for the terminally ill. *Palliat Med* 2000;14:375–385.
- Higginson IJ, Finlay IG, Goodwin DM, Hood K, Edwards AG, Cook A, Douglas HR, Normand CE: Is there evidence that palliative care teams alter end-of-life experiences of patients and their caregivers? *J Pain Symptom Manage* 2003;25:150–168.
- Homsí J, Walsh D, Nelson KA, LeGrand SB, Davis M, Khawam E, Nouneh C: The impact of a palliative medicine consultation service in medical oncology. *Support Care Cancer* 2002;10:337–342.
- O'Mahony S, Blank AE, Zallman L, Selwyn PA: The benefits of a hospital-based inpatient palliative care consultation service: Preliminary outcome data. *J Palliat Med* 2005;8:1033–1039.
- Rabow MW, Dibble SL, Pantilat SZ, McPhee SJ: The comprehensive care team: A controlled trial of outpatient palliative medicine consultation. *Arch Intern Med* 2004;164:83–91.
- Burns JP, Mello MM, Studdert DM, Puopolo AL, Truog RD, Brennan TA: Results of a clinical trial on care improvement for the critically ill. *Crit Care Med* 2003;31:2107–2117.
- Lilly CM, De Meo DL, Sonna LA, Haley KJ, Massaro AF, Wallace RF, Cody S: An intensive communication intervention for the critically ill. *Am J Med* 2000;109:469–475.
- Stapleton RD, Engelberg RA, Wenrich MD, Goss CH, Curtis JR: Clinician statements and family satisfaction with family conferences in the intensive care unit. *Crit Care Med* 2006;34:1679–1685.
- Curtis JR, Treece PD, Nielsen EL, Downey L, Shannon SE, Braungardt T, Owens D, Steinberg KP, Engelberg RA: Integrating palliative and critical care: Evaluation of a quality-improvement intervention. *Am J Respir Crit Care Med* 2008;178:269–275.
- McDonagh JR, Elliott TB, Engelberg RA, Treece PD, Shannon SE, Rubenfeld GD, Patrick DL, Curtis JR: Family satisfaction with family conferences about end-of-life care in the intensive care unit: Increased proportion of family speech is associated with increased satisfaction. *Crit Care Med* 2004;32:1484–1488.
- Campbell DE, Lynn J, Louis TA, Shugarman LR: Medicare program expenditures associated with hospice use. *Ann Intern Med* 2004;140:269–277.
- Gozalo PL, Miller SC, Intrator O, Barber JP, Mor V: Hospice effect on government expenditures among nursing home residents. *Health Serv Res* 2008;43(1 Pt 1):134–153.
- Taylor DH, Jr., Ostermann J, Van Houtven CH, Tulskey JA, Steinhauer K: What length of hospice use maximizes reduction in medical expenditures near death in the US Medicare program? *Soc Sci Med* 2007;65:1466–1478.
- Morrison RS, Penrod JD, Cassel JB: Caust-Ellenbogen M, Litke A, Spragens L, Meier DE; Palliative Care Leadership Centers' Outcomes Group: Cost savings associated with US hospital palliative care consultation programs. *Arch Intern Med* 2008;168:1783–1790.
- Penrod JD, Deb P, Luhrs C, Dellenbaugh C, Zhu CW, Hochman T, Maciejewski ML, Granieri E, Morrison RS: Cost and utilization outcomes of patients receiving hospital-based palliative care consultation. *J Palliat Med* 2006;9:855–860.

20. Back AL, Li YF, Sales AE: Impact of palliative care case management on resource use by patients dying of cancer at a Veterans Affairs medical center. *J Palliat Med* 2005;8:26–35.
21. Maciejewski ML, Diehr P, Smith MA, Hebert P: Common methodological terms in health services research and their synonyms. *Med Care* 2002;40:477–484.
22. Goldstein N, Deb P, Penrod JD: When and how to consider an instrumental variable in palliative care research. *J Palliat Med* (in press).
23. Haro JM, Kontodimas S, Negrin MA, Ratcliffe M, Suarez D, Windmeijer F: Methodological aspects in the assessment of treatment effects in observational health outcomes studies. *Appl Health Econ Health Policy* 2006;5:11–25.
24. Wunsch H, Linde-Zwirble WT, Angus DC: Methods to adjust for bias and confounding in critical care health services research involving observational data. *J Crit Care* 2006;21:1–7.
25. Posner MA, Ash AS, Freund KM, Moskowitz MA, Shwartz M: Comparing standard regression, propensity score matching, and instrumental variables methods for determining the influence of mammography on stage of diagnosis. *Health Serv Outcomes Res Methodol* 2001;2:279–290.
26. D'Agostino RB: Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–2281.
27. Rosenbaum PR, Rubin DB: The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
28. Rosenbaum PR, Rubin DB: Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 1985;39:33–38.
29. Rubin DB: Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127(8 Pt 2):757–763.
30. Christakis NA, Iwashyna TJ: The health impact of health care on families: A matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses. *Soc Sci Med* 2003;57:465–475.
31. Apolone G, Bertetto O, Caraceni A, Corli O, De Conno F, Labianca R, Maltoni M, Nicora M, Torri V, Zucco F; Cancer Pain Outcome Research Study Group: Pain in cancer. An outcome research project to evaluate the epidemiology, the quality and the effects of pain treatment in cancer patients. *Health Qual Life Outcomes* 2006;4:7.
32. Gozalo PL, Miller SC: Hospice enrollment and evaluation of its causal effect on hospitalization of dying nursing home patients. *Health Serv Res* 2007;42:587–610.
33. Tu MS, Chiou CP: Perceptual consistency of pain and quality of life between hospice cancer patients and family caregivers: A pilot study. *Int J Clin Pract* 2007;61:1686–1691.
34. Casarett D, Pickard A, Bailey FA, Ritchie C, Furman C, Rosenfeld K, Shreve S, Chen Z, Shea JA: Do palliative consultations improve patient outcomes? *J Am Geriatr Soc* 2008;56:593–599.
35. Maida V, Ennis M, Irani S, Corbo M, Dolzhykov M: Adjunctive nabilone in cancer pain and symptom management: A prospective observational study using propensity scoring. *J Support Oncol* 2008;6:119–124.
36. Taylor DH, Jr., Steinhilber K, Tulskey JA, Rattliff J, Van Houtven CH: Characterizing hospice discharge patterns in a nationally representative sample of the elderly, 1993–2000. *Am J Hosp Palliat Care* 2008;25:9–15.
37. Shah BR, Laupacis A, Hux JE, Austin PC: Propensity score methods gave similar results to traditional regression modeling in observational studies: A systematic review. *J Clin Epidemiol* 2005;58:550–559.
38. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S: A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* 2006;59:437–447.
39. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V: Principles for modeling propensity scores in medical research: A systematic literature review. *Pharmacoepidemiol Drug Saf* 2004;13:841–853.
40. Baser O: Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health* 2006;9:377–385.
41. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BH, Leufkens HG, de Boer A: Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol* 2004;57:1223–1231.
42. Rosenbaum PR, Rubin DB: Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984;79:516–524.

Address reprint requests to:

Helene Starks, Ph.D., M.P.H.

University of Washington

Department of Bioethics and Humanities

Box 357120

Seattle, WA 98195-7120

E-mail: tigiba@u.washington.edu